STATISTICAL ANALYSIS OF SEVERITY OF MOTOR VEHICLE ACCIDENTS IN SRI LANKA

[1] N.A.M.R.Senaviratna, [2]T.M.J.A.Cooray

[1]Senior Lecturer, [2]Retired Senior Lecturer
[1]Department of Mathematics,
[1]The Open University of Sri Lanka, Sri Lanka.

Abstract : Increasing road accidents and traffic flow is a heavy burden to a developing country like Sri Lanka. The objective of this study is to identify the significant factors affecting motor vehicle accidents in Sri Lanka. Secondary data used in this study between the period 2014 to 2016 were acquired from the traffic police headquarters in Sri Lanka. A total number of 78531 motor vehicle accidents were included in the analysis. Factors considered in the study were vehicle type, gender of driver, validity of license, accident cause, alcohol test, time of accident, weekday/weekend, road surface, weather condition, light condition, location and age of driver. Two third of data (52354) was used to develop the model, and the remaining 1/3 of data (26177) was used to validate the model. Severity of accidents was categorized as grievous and non-grievous accidents. Chi-square test of independence has detected that road surface and weather are not significantly associated with the severity of accidents. The light condition variable is removed due to multicollinearity. Binary logistic regression is applied to model the severity of road accidents due to the dichotomous nature of the dependent variable. The area under the receiver operating characteristic (ROC) curve was 0.692 which means the fitted model classifies the group significantly better than by chance. The fitted model is correctly predicted 79.9 % of the validation data which is greater than the predictive power of the baseline model 69.8%. Results revealed that vehicle type, validity of license, time of accident, location type, alcohol test, accident cause, age of the driver and gender have a significant effect on the severity of accidents. Moreover, Aggressive or negligent driving, driving on straight road, driving in daytime, driving light vehicles have a high chance to be a grievous accident. These findings can aid in modifying laws and establishing preventive approaches in Sri Lanka.

IndexTerms – **accidents, Logistic Regression, Accident severity, Grievous accidents, Non-grievous accidents**.

I. INTRODUCTION

Road accidents are a leading cause for many of deaths around the world. World Health Organization (WHO) has found more than 1.2 million people die each year on the world's roads and most of these deaths are in low and middle-income countries. WHO indicated road traffic injuries are currently estimated to be the 9th leading cause of death across all age groups globally and predicted to become the 7th leading cause of death by 2030. Road accidents are highly influenced to the public health in a country. Furthermore, increasing road accidents evolve social and economic problems due to loss of lives and damage properties.

Ever increasing road accidents and traffic flow is a heavy burden to a developing country like Sri Lanka. The rate of increase in road accidents is 7% per year in Sri Lanka. Increasing in vehicle population is 11% per year. The analysis of past accident data has clearly shown that in Sri Lanka about 50,000 accidents occur annually on average out of which 2000 were fatal accidents and 15,000 were injury accidents. Traffic Police reveals that a Sri Lankan is killed in a road accident every three and half hours and two are critically injured. This is a heavy economic burden to the country.

The objectives of this study are to identify the significant factors affecting motor vehicle accidents in Sri Lanka and estimating the effect of the statistically significant factors on accident severity.

II. LITERATURE REVIEW

According to Renuraj, et al., 2015, type of vehicle and age are the influential variables for severity of road accidents. Another study conducted in Sri Lanka during the period 2010-2014 revealed that variables such as light condition, age of the driver, the validity of the license, urban / rural, weather, vehicle type and age of the vehicle have a decreasing effect on the probability of a fatal accidents. Location type, alcohol test and accident cause have an increasing effect on the probability of a fatal accident. Among them, Accident Cause is the most important variable in the model (Dhananjaya & Alibuhtto, 2016). Furthermore, Liyanage & Rengarasu, 2015 found that experience of the driver (year of driver license issue), vehicle type, light condition and time of the accident are the significant factors for road accidents. Senasinghe, Wirasinghe, & De Barros, 2017 found that age, gender, protection, light conditions, urbanicity, traffic control, and mode of transport, have a significant and direct impact on the severity level of the accidents that occurred on highways.

A study conducted in Riyadh expressed that location and cause of accident were significantly associated with severity of accidents (Al-Ghamdi, 2002). A study done by Wedagama & Dissanayake, 2009 in Indonesia found that probabilities of female motorists contributed more on motor vehicle fatal accidents than males. In addition, age was also significant to influence all vehicle fatalities. A study conducted Chengye & Ranjitkar, 2013 in Auckland motorway revealed that segment length, annual average daily traffic per lane and number of lanes have the most profound effects on accident frequency. According to Chen, et al., 2016, location, vertical alignment, roadside safety rating, driver distraction and overloading of cargo were significant factors for crash severity. In addition, they indicated that intersections were more likely to have side impact on serious road traffic, especially with poor visibility at night. Another study conducted in China revealed that light condition, overloading and gender of driver were highly influential factors on severity of accidents (Zhang, et al., 2013). According to Celik & Oktay, drivers over the age of 65, primary educated drivers, accidents occurring on state routes, highways or provincial roads and the presence of pedestrian crosswalks increase the probability of fatal injuries. The results also indicate that accidents involving cars or private vehicles or those occurring during the evening peak, under clear weather conditions, on local city streets or in the presence of traffic lights decrease the probability of fatal injuries.

III. MATERIALS AND METHODS

3.1. Binary logistic regression

Binary logistic regression model estimates the probability of occurrence of an event by fitting data to a logistic curve. The dependent variable is the population proportion or probability that the resulting outcome is equal to 1. Parameters obtained for the explanatory variables can be used to estimate odds ratios for each of the explanatory variables in the model.

The specific form of the logistic regression model is:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_p x_p}} \tag{1}$$

where π is the probability of the outcome of interest or event, β_0 is the intercept, $\beta_1, \ldots, \beta_p$ are regression coefficients, $x_1, x_2, \ldots, x_p$ are independent variables.

The transformation of the conditional mean $\pi(x)$ logistic function is known as the logit transformation:

$$ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{2}$$

The importance of the logit transformation is that it is linear in its parameters and may range from. $-\infty$ to $+\infty$

There are four assumptions under binary logistic regression. These assumptions are required to satisfy to give a valid result.

- Linearity: The assumption of linearity in logistic regression is that any explanatory variables have a linear relationship with the logit of the response variable. If the relationship between the log odds of the response occurring and each of the explanatory variables is not linear then the model will not be accurate.
- Independent errors: the assumption of independent errors states that errors should not be correlated for two observations.
- Multicollinearity: The assumption requires that explanatory variables should not be highly correlated with each other.
- There should be no outliers, high leverage values or highly influential points.

3.2. Source of Data

Secondary data used from 2014 to 2016 time period in this study were acquired from the police traffic headquarters, Colombo in Sri Lanka. The initial database had 78531 accidents and it was divided into two portions; 2/3 of data (52354) was used to develop the model, and the remaining 1/3 of data (26177) was used to validate the model. The independent variables were vehicle type, gender, validity of the license, age of the driver, accident cause, time, weekday/weekend, road surface, weather condition, light condition, location and alcoholic test result. Response variable is accident severity which consists of two levels namely grievous and non-grievous. Accidents result in death or critically injured are named as grievous accidents and accidents result in non-critically injured or damage only accidents are named as non-grievous accidents.

3.3. Research Methodology

Data analyses are arrayed mainly under preliminary and fundamental analyses. Under preliminary analysis, descriptive statistics was performed to get a general understanding of the whole dataset. In fundamental analysis, Pearson Chi-Square test is performed to check the association between each contributory factor and the severity of accidents. Then multicollinearity is checked and highly correlated variables were removed from the analysis. Finally, due to the dichotomous nature of the dependent variable, binary logistic regression analysis is carried out as advanced analysis to investigate the effect of the variables. These statistical data analyses are conducted by using EVIEWS and SPSS software.

IV. RESULTS AND DISCUSSION

4.1. Descriptive Statistics

Table 1 shows the distribution of factors considered in this study. It can be seen that majority of accidents were occurred in roads with dry surface, clear weather under daylight. The cause of most of accidents was aggressive or negligent driving.

Table 1: Description of Factors

Factor	Levels	Non-grievous (%)	Grievous (%)	Total (%)	Abbreviation
Vehicle Type	Light vehicle	40427(76.5%)	12399(23.5%)	52826(100%)	LV
	Heavy vehicle	18827(43.2%)	6878(26.8%)	25705(100%)	HV
Gender	Male	58268(75.4%)	19043(24.6%)	77311(100%)	M
	Female	986(80.8%)	234(19.2%)	1220(100%)	F
Validity of License	With valid license	51954(76.6%)	15905(23.4%)	67859(100%)	WL
	Without valid license	7300(68.4%)	3372(31.6%)	10672(100%)	WOL
Accident Cause	Speeding	7465((72.1%)	2888(27.9%)	10353(100%)	Cause1
	Aggressive/negligent driving	44462(75.9%)	14070(24.1%)	58532(100%)	Cause2
	Influenced by alcohol/drugs	3055(78.1%)	856(21.9%)	3911(100%)	Cause3
	Others	4272(74.5%)	1463(25.5%)	5735(100%)	Others
Alcohol Test	No alcohol/below legal limit	55677(75.3%)	18267(24.7%)	73944(100%)	BL

	Over legal limit	3577(77.9%)	1010(22.1%)	4587(100%)	OL
Time	Day time	39681(76.4%)	12258(23.6%)	51939(100%)	DT
	Night time	19573(73.6%)	7019(26.4%)	26592(100%)	NT
Weekday/Weekend	Weekday	42519(76.2%)	13242(23.8%)	55761(100%)	WD
	Weekend	16735(73.5%)	6035(26.5%)	22770(100%)	WE
Road surface	Dry	56012(75.5%)	18160(24.5%)	74172(100%)	D
	Wet	3242(74.4%)	1117(25.6%)	4359(100%)	W
Weather Condition	Clear	56067(75.5%)	18175(14.5%)	74242(100%)	CL
	Rainy	3187(74.3%)	1102(25.6%)	4289(100%)	RA
Light condition	Daylight	39070(76.4%)	12069(23.6%)	51139(100%)	DL
	Night, no street lighting	13874(71.6%)	5511(28.4%)	19385(100%)	NSL
	Others	6310(78.8%)	1697(21.2%)	8007(100%)	DD
Location	Bend/Junction	15502(77.8%)	4417(22.2%)	19919(100%)	BJ
	Road	43752(74.6%)	14860(25.4%)	58612(100%)	RD

Then Pearson Chi-square test is performed to test the significant relationship between the independent variables and the dependent variable. According to this analysis, following table describes the as2sociation between each factor and the accident severity.

Table 2: Results of Chi-square test

Variables	**χ^2 value**	**P value**
Vehicle Type	χ^2 (1) = 53.446	0.000
Gender of Driver	χ^2 (1) = 19.271	0.000
Validity of License	χ^2 (1) = 331.396	0.000
Alcohol Test	χ^2 (1) = 16.812	0.000
Accident Cause	χ^2 (2) = 72.183	0.000
Time	χ^2 (1) = 74.149	0.000
Weekday/Weekend	χ^2 (1) = 66.322	0.000
Road Surface	**χ^2 (1) = 2.896**	**0.089**
Weather	**χ^2 (1) = 3.221**	**0.073**
Light Condition	χ^2 (2) = 231.087	0.000
Location	χ^2 (1) = 81.086	0.000

According to the results of Table 2, it can be identified that vehicle type, gender of driver, validity of license, alcohol test, accident cause, time, weekday/weekend, light condition and location type are significantly associated with the accident severity. Only two factors namely road surface and weather, are not significantly associated with the accident severity. Therefore, non-significant factors are removed and continued the analysis.
One of the assumptions in logistic regression is that explanatory variables should not be highly correlated with each other. Therefore, before applying logistic regression, multicollinearity is checked among explanatory variables.

4.2. Checking Correlation between explanatory variables
The correlation coefficients among the explanatory variables can be used as first step to identify the presence of multicollinearity Correlation matrix of highly correlated explanatory variables presented in Table 3. If Pearson correlation coefficient is greater than 0.8 or 0.9 then multicollinearity is a serious concern. Results of Table 3 indicates that the Pearson correlation coefficients between two variables light condition and time are highly correlated and indicated them as bold. These high correlation coefficients signify the presence of severe multicollinearity between the explanatory variable light condition and time of accident.

Table 3: Pearson Correlation matrix between 2 explanatory variables

Variables		Time		Light condition
		DT	NT	NSL
Light condition	DL	**0.978** (0.000)	**-0.978** (0.000)	**-0.782** (0.000)
	NSL	**-0.800** (0.000)	**0.800** (0.000)	1.000 (0.000)

Cell value: correlation coefficient
p value

In addition, tolerance and VIF values were also used to confirm multicollinearity and results were indicated in Table 4.

Table 4: Collinearity statistics

Factor	Levels	Collinearity Statistics	
		Tolerance	VIF
Vehicle type	LV	.961	1.029
	HV	.966	1.035
Validity of license	WL	.966	1.030
	WOL	.964	1.038
Alcohol test	BL	.678	1.475
	OL	.680	1.473
Time	**DT**	**.041**	**24.349**
	NT	**.045**	**21.456**
Weekend/Weekday	WE	.998	1.003
	WD	.993	1.007
Location	RD	.992	1.008
	BJ	.995	1.005
Gender	M	.991	1.010
	F	.989	1.008
Accident cause	Cause1	.976	1.025
	Cause2	.965	1.039
	Cause3	.969	1.031
	Others	.676	1.479
Light condition	**DL**	**.047**	**21.278**
	NSL	**.058**	**20.146**
	Others	**.110**	**9.115**
Age	Age	.989	1.011

Results in Table 4 indicates very low tolerances and larger VIF values for the variables time and light condition. Using these collinearity statistics, it can be concluded that the data almost certainly indicates a serious collinearity problem. Thus, time is removed first from the data and repeats the analysis. However, collinearity still exists among the levels of light variable. Then time is added, and light condition is removed and repeats the analysis. Results are presented in Table 5.

Table 5: Collinearity statistics for remained variables

Factor	Model	Collinearity Statistics	
		Tolerance	VIF
Vehicle type	LV	.961	1.029
	HV	.967	1.035
Validity of license	WL	.995	1.006
	WOL	.965	1.036
Alcohol test	BL	.678	1.475
	OL	.677	1.471
Time	DT	.970	1.031
	NT	.965	1.029
Weekend/Weekday	WE	.997	1.002
	WD	.993	1.002
Location	RD	.997	1.003
	BJ	.992	1.007
Gender	M	.991	1.009
	F	.990	1.011
Accident cause	Cause1	.976	1.024
	Cause2	.981	1.031
	Cause3	.968	1.030
	Others	.676	1.479
Age	Age	.989	1.011

According to results in Table 5, tolerances for all the predictors are very close to 1 and all the VIF values are smaller than 2.5. Therefore, it can be concluded that multicollinearity is not a concern when one of the correlated variables is omitted.

4.3. Binary logistic regression analysis

Since the response variable is dichotomous (grievous/non-grievous), the binary logistic regression model is applied for the data. The maximum likelihood procedure is used to estimate the parameters of the logistic regression model. Forward stepwise selection method was applied under the binomial logistic regression analysis, with variable entry testing based on the significance of the score statistic (the significance level was set at $p<0.05$), and removal testing based on the probability of a likelihood ratio statistic based on the maximum likelihood estimates (the significance level was set at $p>0.10$).

4.3.1. Baseline Model

Table 6 presents the results of the baseline model which is the model with only the constant included before explanatory variables are entered into the model. Logistic regression compares this model with a model including all the significant factors to determine whether the latter model is more appropriate.

Table 6: Results of baseline model

Baseline Model	B	S.E.	Wald	df	Sig.	Exp(B)
Constant	1.162	.010	12828.477	1	.000	3.196

Initial -2 Log Likelihood: 57499.931

Wald Chi-Square tests the null hypothesis that the constant equals 0. Results of the Table 6 shows that constant is statistically significant to the model. Initial log likelihood value of the baseline model is 57499.931. This value is used to select an optimal model. Predictive power of the baseline model is 69.5%, which indicates the overall percentage of correctly classified cases when there are no explanatory variables in the model.

4.3.2. Developed model interpretation

Predictors with positive coefficients cause an increasing tendency to result into fatalities. Similarly, negative coefficients indicate decreasing tendency for those significant predictors. Table 7 explains the variables in the developed model used to predict the severity of accidents.

Table 7: Variables in the model

Variables	B	S.E.	Wald	df	Sig.	Odd ratio
Heavy vehicle	-.183	.022	68.737	1	.000	.833
Without license	-.444	.029	15.826	1	.000	.641
Night time	-.143	.022	43.081	1	.000	.866
Road	.175	.024	51.573	1	.000	1.191
No alcohol /below limit	.301	.056	28.445	1	.000	1.351
Aggressive/negligent driving	.120	.030	236.856	1	.000	1.128
Age	-.138	.023	37.564	1	.000	.871
Male	.201	.094	4.514	1	.034	1.222
Constant	1.219	.032	1479.642	1	.000	3.382

When exploring results of the Table 7, validity of license, vehicle type, location type, time, age of driver, alcohol test, accident cause and gender have a significant effect on the severity of accidents. Odds ratio of vehicle type indicates that a grievous accident occurred by heavy vehicles is 83% less likely to be a grievous accident occurred by light vehicles. Similarly, a grievous accident occurred by drivers who, without a valid license are 64% less likely to be odds of a grievous accident occurred by drivers who having a valid license. Odds of a grievous accident occurred in nighttime is 87% less likely to be a grievous accident occurred in daytime. Odds of a grievous accident occurred in road is 19% more likely to be a grievous accident occurred in bend/junction. Odds of a grievous accident occurred by drivers who used alcohol below legal limit or no alcohol 35% more likely to be a grievous accident occurred by drivers who used alcohol over legal limit. Odds of a grievous accident occurred by aggressive/negligent driving is 13% more likely to be a grievous accident occurred by speeding. The odds ratio of age is 0.871. It indicates that for everyone unit increase in age (one additional year of living), the odds of occurring a grievous accident decreases which implies the older the driver, the less the accident risk. Odds of a grievous accident occurred by male drivers 22% more likely to be a grievous accident occurred by female.

4.3.3. Logit Model

The logit model with the significant variables is:

$$logit(p) = 1.219 - 0.183HV - 0.444WOL - 0.143NT + 0.175BJ + 0.301BL + 0.12Cause2 - 0.138Age + 0.201M$$

4.3.4. Importance of Variables in the Model

Table 8 presents the information how the model is affected if an explanatory variable is added to the model. In other words, which variable is important for the model. Results of following table are used to examine the importance of a variable in the model.

Table 8: Importance of variables in the model

Step	Improvement			Model			Variable
	Chi-square	df	Sig.	Chi-square	df	Sig.	
1	201.658	1	.000	201.658	1	.000	IN: Accident cause
2	65.982	1	.000	267.640	2	.000	IN: Vehicle Type
3	55.460	1	.000	323.100	3	.000	IN: Location
4	39.785	1	.000	362.885	4	.000	IN: Time
5	36.441	1	.000	399.326	5	.000	IN: Age
6	32.678	1	.000	432.004	6	.000	IN: Alcohol
7	16.706	2	.000	448.710	8	.000	IN: License
8	4.221	1	.030	452.931	9	.000	IN: Gender

Table 8 indicates that adding the variable accident cause to the model makes the biggest change in the model's log likelihood value. Therefore, accident cause is the most important variable in this model. It is followed by the vehicle type, location type, time, age of driver, alcohol test, validity of license and gender respectively.

4.4. Measures of Goodness of Fit

4.4.1. Test of Model coefficients

Table 9: Test of developed model coefficients

	Chi-square	df	p value
Step 8	4.221	1	0.028
Block	452.931	9	0.000
Model	452.931	9	0.000

Results of Table 9 indicates that the chi-square is highly significant (chi-square=452.931, p<0.000 with df =9). Thus, it can be concluded that the developed model is significantly better than the baseline model. That means the accuracy of the model improved when added the explanatory variables.

4.4.2. Model Summary

Table 10: Model Summary

Step	-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
8	57047.735	.604	.651

According to the results of the Table 10, the developed model has a significantly reduced log likelihood value (57047.735) compared to the baseline model. It is revealed that the developed model is explaining more of the variance in the outcome and it is an improvement over the baseline model. Thus, it can be concluded that the developed model is better at predicting the severity of the accidents than the baseline model where no predictor variables were added. According to Cox & Snell R^2 and Nagelkerke R^2, the explained variation in the dependent variable based on the model are 60.4% and 65.1% respectively.

4.4.3. Predictive Accuracy of Developed Model

Table 11: Classification Table

Observed	Predicted		
	Grievous	Non-grievous	Percentage Correct
Grievous	9123	3355	73.11
Non-grievous	8953	30911	77.54
Overall Percentage			76.47

Table 11 indicates that 73.1% were correctly classified for grievous accidents and 77.5% for non-grievous accidents. Overall, 76.5% were correctly classified. It can be seen that the developed model is correctly classifying the outcome for 76.5% of the cases compared to 69.5% in the null model.

4.4.4. Hosmer and Lemeshow test

Table 12: Results of Hosmer and Lemeshow test

Chi-square	df	Sig.
7.755	8	0.458

As the results shown in the Table 12, Hosmer & Lemeshow test of the goodness of fit suggests the model is a good fit to the data as p=0.458 (>0.05).

4.4.5. Receiver Operating Characteristic Curve

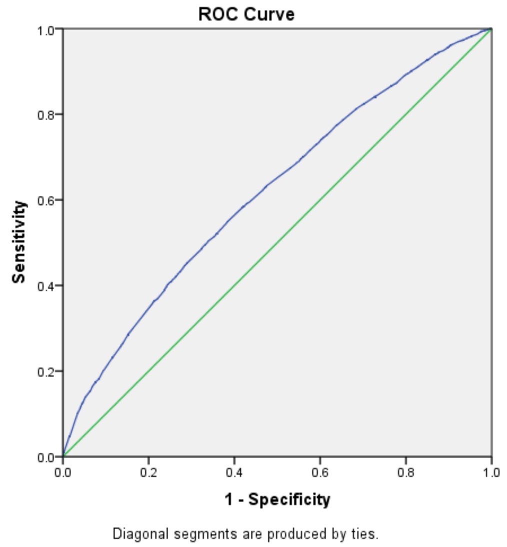

Figure 1: ROC Curve

Table 13: Area Under the Curve

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.692	.003	.000	.687	.697

According to the above Table and Figure, the area under the curve is 0.692 with 95% confidence interval (0.687, 0.697). Moreover, the area under the curve is significantly different from 0.05 since the p =0.000 <0.05. That means, the logistic regression classifies the group significantly better than by chance.

4.4.6. Detecting Influential Observations

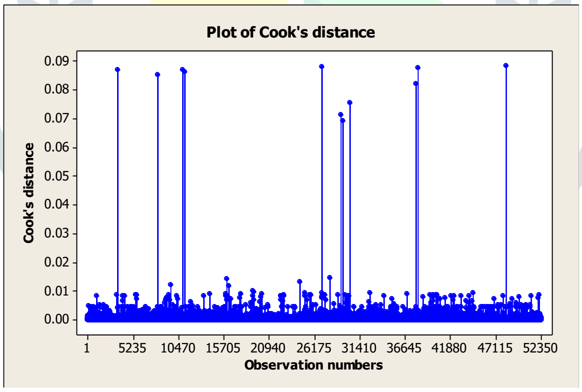

Figure 2: Plot of Cook's Distance

It can be seen that all observations are less than 1 and even less than 0.5. Therefore, it can be said that there are no outliers nor influential observations in this data set.

4.5. Validation of the Model

Table 14: Category prediction

Observed	Predicted		
	Grievous	Non-grievous	Percentage Correct
Grievous	5026	1773	73.9
Non-grievous	3482	15896	82.0
Overall Percentage			79.9

Table 14 indicates that the model correctly predicted 79.9 % of the validation data which is greater than to the predictive power of the baseline model 69.8%. That means the developed model more accurately predicts the severity of accidents than the prediction in baseline model.

V. CONCLUSION

The results discovered that vehicle type, validity of license, time of accident, location type, alcohol test, accident cause, age of the driver and gender have a significant effect on the severity of accidents. However, weekday/weekend variable is not significantly associated with the severity of motor vehicle accidents. Moreover, aggressive/negligent driving, driving on straight road, driving at daytime and driving light vehicle have a high chance to be a grievous accident. In addition, the younger drivers have more accident risk. Finally, it is concluded that accident cause is the most important variable in the model. This is an issue which needs high level attention from drivers and high commitment by traffic police. Drivers have a great responsibility to reduce road accidents and control their ambience.

REFERENCES

[1] Al-Ghamdi, A., (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis and Prevention*, Volume 34, pp. 729-741.

[2] Allison, P., (2014). Measures of fit for logistic regression. *SAS Global Forum* Statistical Horizons LLC, pp 60-72.

[3] Amarasingha, N., & Dissanayake, S (2013). Modelling frequency of truck crashes on limited access highways. Journal of the Transportation Research Forum, 123-139.

[4] Baruah, A. & Chaliha, R., (2015). A study of incidence of alcohol use in fatal road traffic accidents. *Journal of Indian Academy of Forensic Medicine,* 37(1), pp. 12-15.

[5] Bursac, Z., Gauss, C., Williams, D. & Hosmer, D., (2008). Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine*, pp. 3-17.

[6] Celik, A. & Oktay, E., (2014). A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey. *Accident Analysis and Prevention,* pp. 66-77.

[7] Chen, Y. et al., (2016). Differences in factors affecting various crash types with high numbers of fatalities and injuries in China. *PLOS One,* pp. 1-12.

[8] Chengye, P., & Ranjitkar, P. (2013). Modelling motorway accidents using Negative Binomial Regression. Journal of Eastern Asia Society for Transportation Studies,10,1946-1963.

[9] Dhananjaya, S. & Alibuhtto, M., (2016). Factors Influencing Road Accidens in Sri Lanka.: A Logistic Regression Approach. *Annual Science Research Sessions-South Eastern University of Sri Lanka*, pp. 157-173.

[10] Dissanayaka, M. & Kulatunga, D., (2012). Analysis of fatalities in road accidents considering Peliyagoda police area in Gampaha. *Annual Science Research Sessions-University of Kelaniya*, p. 45.

[11] Giancristofaro, R. & Salmaso, L., (2003). Model performance analysis and model validationnin logistic regression. *STATISTICA anno LXIII,* pp. 375-396.

[12] Haadi, A., (2014). Identification of Factors that Cause Severity of Road Accidents in Ghana: a Case Study of the Northern Region. *International Journal of Applied Science and Technology,* May, pp. 242-249.

[13] Hosmer, D., Hosmer, T., Lemeshow , S. & Cessie , L., (1997). A comparison of goodness of fit tests for the logistic regression model. *Statistics in medicine,* pp. 965-980.

[14] Jeepara, P. & Pirasath, S., (2011). Road traffic accidents in Eastern Sri Lanka: An analysis of admissions and outcome. *Sri Lanka Journal of Surgery,* 29(2), pp. 72-76.

[15] Komba, D. D., (2006). Risk factors and road traffic accidents in Tanzania: A case study of Kibaha District, *Trondheim: Norwegian University of Science and Technology,* pp 24-33.

[16] Kumarage, A. S., Wickramasingha, S. M. & Jayaratne, M. D., (2003). Analysis of road accidents in Sri Lanka. *Supreme Group of Companies*, pp 56-64.

[17] Liyanage, T. & Rengarasu, T., (2015). Traffic Accident Analysis and Development of Accident Prediction Model.*The Institution of Engineers, Sri Lanka*, pp. 105-109.

[18] Lund, B. & MI, D., (2015). Fitting and evaluating logistic regression models. *Marketing Associates, LLC, pp* 105-115.

[19] Midi, H., Sarkar, S. & Rana, S., (2013). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics,* pp. 253-267.

[20] Mohamed Aslam, A., (2015). An Econometric Analysis of Trafffic Accident in Sri Lanka. pp. 1077-1084.

[21] Rana, S., Midi, H. & Sarkar , S.(2005). Validation and Performance Analysis of Binary Logistic Regression Model. *Proceedings of the WSEAS International Conference on ENVIRONMENT, MEDICINE and HEALTH SCIENCES,* pp. 51-55.

[22] Renuraj, S., Varathan, N. & Satkunananthan, N., (2015). Factors Influencing Traffic Accidents in Jaffna. *Sri Lankan Journal of Applied Statistics,* September, 16(2), pp. 117-133.

[23] Robin, P., (2014). Use on multinomial logistic regression in work zone crash analysis for Missouri work zones. *Missouri University of Science and Technology*, pp 45-61.

[24] Sarkar, S., Midi, H. & Rana, S., (2011). Detection of outliers and influential observations in binary logistic regression: An empirical study. *Journal of Applied Sciences,* pp. 26-35.

[25] Seiss, M., (2009). Logistic and Poisson Regression: Modeling Binary and Count Data. *VirginiaTech.*

[26] Senadeera, W. A., (2016). Analyzing traffic accidents in Gampaha district. *Academic Sessions- University of Sri Jayewardenepura,* pp 67-70.

[27] Senasinghe, A.P., Wirasinghe, S.C., & De Barros, A.G (2017). Road safety issues on two major intercity highways in Sri Lanka (A001 and A004)- Preliminary analysis. R4TLI Conference Proceedings, pp.201-206.

[28] Shruthi, P. et al., (2013). Analysis of fatal road traffic accidents in a Metropolitan City of South India. *Journal of Indian Academy of Forensic Medicine,* 35(4), pp. 317-320.

[29] Singh, D., Moorthi, K., Singh, S. & Goel, S., (2014). Profile of road traffic fatalities in adults: A 40 year study in Chandigarh Zone of North West India. *Journal of Indian Academy of Forensic Medicine,* 36(1), pp. 47-51.

[30] Singh, H. & Aggarwal, A. D., (2010). Fatal road traffic accidents among young children. *Journal of Indian Academy of Forensic Medicine,* 32(4), pp. 286-288.

[31] Singh, R. et al., (2013). Elucidation of risk factors in survivors of road traffic accidents in North India. *Hard Tissue,* 2(1), pp. 1-6.

[32] Somasundaraswaran , A. K., (2006). Accident statistics in Sri Lanka. *International Association of Traffic and Safety Sciences Research,* pp. 115-117.

[33] Tanaboriboon , Y. & Satinnam , T., (2005). Traffic accidents in Thailand. *IATSS Research,* 29(1), pp. 88-100.

[34] Wedagama, D. & Dissanayake, D., (2009). The Influence of Accident Related Factors on Road Fatalities Considering Bali Province in Indonesia as a Case Study. *Journal of the Eastern Asia Society for Transportation Studies*, Volume 8.

[35] Yan, X., Radwan, E. & Abdel-Aty, M., (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accident Analysis and Prevention,* pp. 983-995.

[36] Zhang, G., Yau, K. & Chen, G., (2013). Risk factors associated with traffic violations and accident severity in China. *Accident Analysis and Prevention,* pp. 18-25.

[37] Zhu, D., (2014). Analysis of factors affecting motorcycle-motor vehicle crash characteristics. *Ohio University of Dayton*, pp 123-143.